**A computational perspective on faith: religious reasoning and Bayesian decision**

Francesco Rigoli[1],

[1] *City, University of London, Northampton Square, London, EC1V 0HB, UK*

**Correspondence**: Francesco Rigoli
Department of Psychology
City, University of London
Northampton Square, London, UK EC1V 0HB
francesco.rigoli@city.ac.uk

**Abstract**

Religious reasoning (the processes through which religious beliefs are formed) has been investigated by two different approaches. First, explanation theories portray religious reasoning as arising for *explaining* salient aspects of reality. Second, motivation theories interpret religious reasoning as driven by other motives, for example fostering community bonding. Both approaches have provided fundamental insight, yet whether they can be reconciled remains unclear. To address this, I propose a unifying computational theory of religious reasoning expressed in mathematical terms. Although a mathematical approach has been rarely applied to study religion, its advantage is describing a phenomenon clearly and formally. Relying on a Bayesian decision framework, the model comprises three key elements: prior beliefs, novel evidence, and utility. The first two describe the processes classically interpreted by explanation theories, while utility captures phenomena highlighted by motivation theories. By reconciling explanation and motivation theories, this proposal offers a unifying computational picture of religious reasoning.

## 1. Introduction

Religion is an important phenomenon characterising all known human cultures. Although the definition of this multifaceted concept remains highly controversial, a common perspective conceives religion as encompassing (i) a set of beliefs viewing important aspects of life as influenced by supernatural agents such as gods and spirits, and (ii) a set of practices available to interact with those agents (e.g., Boyer, 2001; Tylor, 1871; Stark & Finke, 2000). A classical question in the scientific study of religion is where religious beliefs come from. The theories that have attempted to answer this question are many, but I suggest grouping them in two families. The first family (explanation theories) includes accounts proposing that religious beliefs arise primarily for *explaining* salient aspects of life and reality (e.g., Andersen, 2019; Barrett, 2000; Boyer, 2001; Guthrie, 1993; Hogg et al., 2010; Iannaccone, 1998; McCauley, 2017; Schjoedt et al., 2013; Stark & Finke, 2000; Taves & Asprem, 2017; Tylor, 1871). Some of these accounts emphasise pure epistemic needs, implying that religion's attempt to explain is an end in itself (e.g., Hogg et al., 2010; Tylor, 1871). Other explanation theories conceive religious beliefs as specific forms of inference produced by the brain when presented with certain types of stimuli (e.g., Barrett, 2000; Boyer, 2001; Guthrie, 1993; McCauley, 2017; Van Leeuwen & van Elk, 2019). Finally, a third group of explanation theories attribute a central role to utilitarian motives such as securing well-being after death (Iannaccone, 1998; Stark & Finke, 2000). According to these views, the final goal of religion is utilitarian and not epistemic, but explanations afforded by religion are still considered as necessary to fulfil utilitarian goals. Hence religious beliefs are still understood as attempts to explain, and utilitarian goals are thought not to bias these beliefs. A point where explanation theories disagree is to what extent religious beliefs (and human beliefs in general) are rational. Some scholars have argued that, after all, common people reason rationally, albeit constrained by their brain's limited capacity (Simon, 1997). Religious reasoning would be no exception (Andersen, 2019; Iannaccone, 1998; Stark & Finke, 2000; Taves & Asprem, 2017). For example, it has been argued that apparently bizarre religious beliefs expressed by "primitive tribes" appear irrational only from the privileged perspective of modern science, but sound clever solutions if looked from the

tribes' internal perspective (Evans-Pritchard, 1937). However, other scholars have highlighted the ubiquity of biases and illogicities in human reasoning, casting doubts on whether human beliefs can be considered rational, or even bounded rational (Shermer, 2002; Wolpert, 1994). According to these scholars, religious beliefs would represent paradigmatic examples of human fundamental irrationality. Another debated issue among explanation theories regards the influence of evolution. When examining how religious beliefs develop, some scholars downplay the role of innate factors and advocate a central role for experience and learning. For example, individuals' beliefs about God are suggested to reflect early experience with parents (Cassibba et al., 2008; Granqvist et al., 2007; 2010). Conversely, other scholars argue that key aspects of religion emerge from how the human brain has been shaped by evolution (e.g., Boyer, 2001; McCauley, 2017). For instance, our innate tendency to interpret certain sensory experiences in terms of agency would be at the root of religion's emphasis on supernatural agents such as gods and spirits (Atran, 2002; Bloom, 2007; Barrett, 2000; Guthrie, 1993).

Contrary to explanation theories, motivation theories postulate that religious beliefs do not derive from an attempt to explain life and reality. Other motives would be critical. To illustrate the difference between explanation and motivation theories, consider a person arbitrating between the hypothesis that existence continues after death versus the hypothesis that existence ceases after death. According to explanation theories, the person will endorse the hypothesis more in line with available evidence. Conversely, according to motivation theories the person will endorse the hypothesis that satisfies motives such as avoiding anxiety, independent of its fit with evidence. In other words, motivation theories postulate that some forms of motivated reasoning underly religious beliefs (Kunda, 1990; Willer, 2009). A multiplicity of views exists regarding which motives might drive the formation of religious beliefs. Following Hume (1757) and Freud (1927), comfort theories propose that religious beliefs develop primarily to sedate anxiety, especially death anxiety (Atran & Norenzayan, 2004; Norenzayan & Hansen, 2006; Swan, 2019; Willer, 2009; Vail et al., 2010). Inspired by Marx and Engels (1945), other theories build on the notion of ideology and view religious beliefs as developed

by the ruling class to promote its interests (Jost et al., 2014; Williams, 1996; Fanon, 1963; Fulton, 1987). Within this tradition, scholars such as Gramsci (1971) have argued that lower classes too can develop religious beliefs in support of their interest, and not only as the result of hegemonic influences exerted by ruling classes (Williams, 1996; Fulton, 1987). A third tradition of motivation theories can be traced back to Kant's idea that religious beliefs arise to legitimise and spread moral rules (Batson et al., 1993; Saroglou, 2011; Vitell et al., 2009). Finally, following early intuitions of Durkheim (1912), some scholars have claimed that religious beliefs are developed to foster bonding and feelings of belonging within a community (Baumeister, 1991; Krause & Wulff, 2005; Pargament et al., 1983).

Explanation and motivation theories have both contributed to shed light on fundamental aspects of religion. However, the co-existence, and even thriving, of such different perspectives represents a conundrum if one aims at understanding religion as a unitary phenomenon (though some remain sceptics about this aim, yet parsimony continues to motivate scholars to search for a unifying framework to understand religion). Is there any possible solution? Can one choose one of the two perspectives? (and within it, can one specific account be selected?). This seems a difficult choice, as each perspective offers explanations for certain aspects which remain hard to interpret by the other perspective. Many scholars today believe the two perspectives ultimately need to be integrated (Atran & Norenzayan, 2004; Swan, 2019). But how? To date, the question of whether the two perspectives can be integrated in a meaningful way remains open. The goal of this paper is to address this question and offer a possible way to reconcile explanation and motivation theories of religion under a unifying framework. To illustrate its principles in a formal and clear way, the framework is described adopting computational modelling. Such computational approach has rarely been applied to the investigation of religion, perhaps because this is often viewed as a phenomenon which is too complex and ineffable. However, I explore the possibility that a computational approach can offer a clear and precise characterisation of key aspects of religion, although surely some of the subtleties will elude this account. Because the mathematical formalism adopted is Bayesian decision (Bishop, 2006), the model described here is referred to as Bayesian Decision Model of Religion (BDMR). The paper is structured

as follows. The next section (section 2) describes the model in detail. Afterwards (section 3), the role of prior beliefs and evidence (two key elements of the model) is examined together with their link with explanation theories. This is followed by a treatment of the role of utility in the model and its relevance for motivation theories (section 4). Next, the model is examined considering empirical research about the psychology of religion (section 5). Finally, the BDMR is discussed with respect to more general issues (section 6).

## 2. The Bayesian decision model of religion

Humans have constantly to choose among different interpretations of the experiences they make. Sometimes religious interpretations for these experiences are available. An interpretation can be defined as religious when a phenomenon is viewed as dependent on the power or action of spirits or gods. Sometimes a religious interpretation might compete with other religious interpretations, other times it might compete with interpretations with no religious content. As an example of two competing religious interpretations, an individual might wonder whether her illness is effectively God's punishment for a bad action, or it is the product of a persecution of an ancestor's spirit. The individual might consider these possibilities against a non-religious explanation which interprets the illness as contagion from a family member. This paper aims at examining how people reason when religious hypotheses are under consideration, a process I refer to as religious reasoning. To explain this process, I pursue a computational approach, in other words I seek to identify the first principles, or the basic meaning, of a psychological phenomenon (i.e., I focus on the computational level of analysis, adopting Marr's famous terminology).

In contemporary psychology and neuroscience, one of the most influential computational perspectives interprets the brain as a Bayesian inference machine (Clark, 2013; Friston & Kiebel, 2009; Knill & Pouget, 2004; Oaksford & Chater, 2007; Rao & Ballard, 1999). The key idea is that, when novel evidence is experienced, the brain integrates new information with prior knowledge in an optimal

fashion. This simple and elegant idea has contributed to understand a variety of phenomena across several domains, from perception to social interaction. Recently, this idea has been fruitfully extended to the study of religion (Andersen, 2019; Schjoedt et al., 2013; Taves & Asprem, 2017; Van Elk et al., 2016), specifically adopting a predictive coding framework (Clark, 2013; Friston & Kiebel, 2009; Rao & Ballard, 1999). For example, predictive coding has been proposed to explain how prior cultural knowledge leads some individuals to interpret certain experiences in religious terms, namely as manifestations of spirits or gods (Taves & Asprem, 2017). I argue that a Bayesian inference approach (and predictive coding) is well-equipped for interpreting explanation theories of religion. This is because the latter theories and Bayesian inference both are built upon assuming a motivation for being accurate. However, Bayesian inference (and predictive coding) seems inadequate to interpret motivation theories, because the latter highlight motives which go beyond accuracy seeking. In order to account for motivation theories, I propose to extend Bayesian inference (and predictive coding) to the notion of Bayesian decision (Bishop, 2006). The idea of the latter is that *utility maximization*, and not *accuracy maximization*, is the ultimate principle guiding religious reasoning. In this framework, being accurate is just a means for increasing utility (though accuracy remains important, otherwise reward is less likely to be collected).

The model I propose corresponds to a standard Bayesian decision framework which I have implemented adopting the formalism of Bayesian networks (Bishop, 2006). The network is represented graphically in Fig. 1 (a more formal description is offered in the Appendix). It describes the beliefs an agent entertains about certain important variables that are relevant for religious interpretations. The variables included in the model are represented by rectangles (for categorical variables) and circles (for continuous variable). Arrows indicate probabilistic dependencies among variables. The first variable in the model is Hypothesis (Hyp), representing a categorical variable reflecting mutually exclusive claims, some of which include religious interpretations. For example, one claim might be that an illness is God's punishment for recent misbehaviour (a religious hypothesis), and the alternative claim that an illness is due to a frequent interaction with an infected patient (a

non-religious hypothesis). Note that each claim can include several statements, provided that ultimately the claims remain mutually exclusive. For example, one claim might be that an illness is due to a frequent interaction with an infected patient *combined* with God's punishment for recent misbehaviour. The alternative statement might be that an illness is due to a frequent interaction with an infected patient *combined* with a weak immune system. Note that the two hypotheses are still mutually exclusive. The first hypothesis can be treated as religious, because it includes at least one religious statement. The second hypothesis can be treated as non-religious, because it includes no religious statements. Hyp plays the central role within the BDMR, because the final result of the model is arbitrating among the different hypotheses implemented by Hyp. The second variable in the model is Prior Belief System (PBS). This represents a categorical variable reflecting a set of more general alternative views on the world, personal life, and society. For example, one view might be that God often intervenes in people's life to guide their behaviour, and the alternative view that God is usually uninterested in mundane affairs. The variable Hyp depends on PBS, as the arrow going from the latter to the former indicates. For example, someone tending to view God as interventionist (PBS) will also tend to attribute higher likelihood to the hypothesis that the illness reflects God's punishment (Hyp).

In the model, both Hyp and PBS are treated as *hidden* (or *latent*) variables, as they cannot be observed directly but need to be inferred indirectly. For example, one does not know for sure whether God tends to be interventionist or not (PBS), nor whether the illness is God's punishment or not (Hyp). In addition to these two hidden variables (Hyp and PBS), the model includes two variables that are directly observed: a direct evidence (DirE) and a social evidence (SocE). These two variables are believed to be the consequence of Hyp, as indicated by arrows going from the latter to DirE and SocE. This probabilistic relation implies that observing the values of the two sensory variables (DirE and SocE) helps inferring the values of the two hidden variables (Hyp and PBS), as will be explained below.

Note that in the BDMR the distinction between DirE and SocE is important, because it emphasizes the different roles played by different sources of information. DirE concerns direct evidence conveyed via

own sensory experience. For example, one may consider a dream about God expressing disappointment as sensory evidence relevant for inferring whether the illness is God's punishment or not. SocE concerns indirect information conveyed by other individuals. This captures the fact that, for humans, others' opinions are critical for informing own opinions, especially in religious matters. Think to the knowledge we acquire via word of mouth, via books, or via other media (also non-verbal communication may be conceived as social evidence). In my example, a trusted family member may express an opinion about the cause of the illness (this advice is reflected in SocE). Both DirE and SocE are represented by continuous variables. In my example, negative values for DirE or SocE correspond to evidence supporting the religious hypothesis (i.e., the illness is God's punishment). Importantly, each evidence variable (DirE and SocE) is associated with a weight (formally, a precision parameter; see Appendix) which determines how influential that evidence is during inference.

Finally, the BDMR includes a Hypothesis Decision (HDec) variable and an Expected Outcome (EOut) variable. HDec is categorical and indicates which hypothesis of the variable Hyp is accepted as true and is used to guide behaviour. For example, HDec may include the following two categories: (i) accept the religious hypothesis (and spend time praying; assuming that praying can help winning God's favour) and (ii) accept the non-religious hypothesis (and do not pray, since praying is time consuming). EOut reflects the expected outcome of this decision and depends both on Hyp and HDec. EOut is represented by a continuous variable where negative values correspond to punishment and positive values to reward. For example, EOut describes the outcome expected to occur (i) if the religious hypothesis is true and I accept it (and spend time praying), (ii) if the non-religious hypothesis is true and I accept it (and do not pray), (iii) if the religious hypothesis is false but I accept it (and spend time praying) (iv) if the non-religious hypothesis is false but I accept it (and do not pray).

The BDMR realizes Bayesian decision by following a sequence of inference steps and eventually deciding which hypothesis to accept. Specifically, the model infers the consequences (in terms of reward or punishment) of accepting different hypotheses considering evidence from DirE and SocE.

Eventually, the hypothesis associated with the best consequence is accepted. More formally, this inference and decision process works as follows. DirE and SocE are observed and inference follows multiple step. At each step, one different category of HDec is considered as observed and the posterior probability of EOut given DirE, SocE and HDec (i.e., P(EOut|DirE, SocE, HDec)) is calculated. This is repeated for all possible categories of HDec. After inference, decision follows, whereby the category of HDec associated with the best EOut (i.e., the highest posterior utility value) is chosen (note that, in a Bayesian framework, choice of one hypothesis does not entail that remaining hypotheses are fully eliminated; all hypotheses remain available, each associated with a "strength" value, formally corresponding to the posterior utility value).

It is important to highlight that, in the BDMR, the selected hypothesis is not necessarily the best supported by evidence (i.e., the one that maximizes accuracy), but the one associated with the best consequences (i.e., the one that maximizes expected utility). This emphasis on utility maximization distinguishes Bayesian decision theory from Bayesian inference. For example, the model predicts that an individual will be more likely to endorse the religious hypothesis if rejecting this hypothesis is perceived as too risky if the hypothesis is eventually true. Based on this reasoning, the model predicts that, when one is more frightened by the illness, the religious hypothesis will be more likely to be endorsed, because rejecting the possibility that the illness is God's punishment (and not praying for receiving God's help to heal) will be evaluated as too risky if this hypothesis is actually true.

However, note that accuracy is still fundamental in the BDMR. This is because accepting a hypothesis which is poorly supported by prior beliefs (PBS) and by evidence (DirE and SocE) is scarcely rewarding, implying that such hypothesis will be discarded. By integrating accuracy and utility drives during religious reasoning, the BDMR offers a principled solution for reconciling explanation and motivation theories of religion, respectively (see below).

According to the BDMR, what is the phenomenological implication of accepting one hypothesis over the other? I propose that the implication is that, phenomenologically, an agent will believe that the

accepted hypothesis is true even if, as explained above, it does not necessarily enjoy more support from evidence. In other words, the BDMR postulates that agents are blind to the inference/decision process described above; they simply perceive the accepted hypothesis as true, without being aware that their perception is the product of utility maximization. In other words, the model assumes a form of motivated reasoning (Kunda, 1990; Willer, 2009) or self-deception during belief formation. Why should this occur? Following Trivers (2011), in an evolutionary perspective beliefs can be understood as having a fundamental pragmatic nature in as much as they enable one to achieve goals. To be effective, beliefs would need to satisfy three fundamental requisites. First, they would need to describe the world with some accuracy, an aspect the BDMR captures by attributing importance to evidence and prior beliefs (if these are ignored, goals will rarely be obtained). Second, they would need to take utility into account, also in line with the BDMR. Third, because the human species has adapted through coordinating complex social behaviour, beliefs will need to persuade others. Only if this occurs, beliefs will ultimately be effective. In this perspective, self-deception during reasoning might have evolved as an effective strategy to persuade others (a possibility which has received empirical support; Smith et al., 2017; Schwardmann & Van der Weele, 2019).

In short, the BDMPR explains religious reasoning by relying on a Bayesian decision framework. This proposes that individuals consider prior belief systems together with novel evidence to infer the consequences of accepting alternative hypotheses, eventually endorsing the hypothesis associated with the highest utility. This inference/decision process is postulated to be subconscious, and to ultimately result in the perception that the accepted hypothesis is true at the phenomenological level. Below, I will examine the role of each element of the model in the genesis of religious reasoning.

## 3. Prior beliefs, evidence, and explanation theories

The BDMR assigns a pivotal role to prior belief systems, captured by the variable PBS (Fig. 2A). This variable can reflect several forms of prior knowledge that can be grouped in three categories. First

(prior religious beliefs), it can simply describe prior beliefs about supernatural agents. In the example above, PBS represents the prior belief that God tends to intervene in people's life versus the alternative belief that God tends to abstain from intervening.

To introduce the second and third categories of prior beliefs, remember that the BDMR is arbitrating among hypotheses regarding supernatural agents. Considering that direct evidence about supernatural agents is normally scarce, knowledge about agents in the real world can be informative on the characteristics of supernatural agents. In other words, prior knowledge about real agents can provide clues about supernatural agents. Following this reasoning, the second type of prior beliefs (prior interpersonal beliefs) concerns knowledge of interpersonal relationships. These beliefs are analogous to the notion of social script and describe the behaviour expected by others in certain social contexts (Abelson, 1981). Building on the concept of internal working model in attachment theory (Bowlby, 1969), prior interpersonal beliefs can also concern intimate relationships such as with parents or partners. There is empirical support for the idea that religious beliefs are linked with beliefs about relational figures (Cassibba et al., 2008; Granqvist et al., 2007; 2010). For example, a link has been observed between attachment style (reflecting beliefs about parents) and beliefs about God, with distant parents associated with God viewed as distant, and caring parents associated with God viewed as caring (Granqvist et al., 2007). As an example of how this can be implemented in the BDMR, one can assign to PBS the category "intimate relational figures are distant" versus the category "intimate relational figures are caring". Like prior interpersonal beliefs, the third type of prior beliefs (prior social beliefs) also leverage on knowledge about real agents, but now on a broader scale, namely focusing on how society and politics are organized. In other words, these prior beliefs suggest that the "supernatural" society (encompassing the gods, spirits and their relationships) is organized like the human society. Durkheim first suggested that religious views reflect the structure of society (1912). He noted that societies organized in clans often develop totemic religions based on spirits symbolized by animals, each connected to one clan. As another example, many antiquity cultures such as the Greeks and the Mesopotamians were organized in independent, and often competing, city states.

These cultures developed a polytheistic religion where multiple gods were acknowledged and where each city established a particular devotion for one deity (e.g., Athens for Athena and Babylon for Marduk). This perspective also raises the possibility that universal monotheistic religions such as Christianity and Islam initially benefited from large multinational empires (the Roman and the Arab empire, respectively). Living in a vast empire where ethnic differences were politically unimportant might have supported the belief in one and universal God (though certainly this was one of many factors, and surely it is not sufficient for the development of monotheism - as studying other ancient empires and other civilizations indicates). To represent this last example using the BDMR, one can assign to PBS the category "the emperor (or calif) is the main political authority" versus the category "the city elderly council is the main political authority". The former category would support the hypothesis (expressed in Hyp) "There is only one God" and the latter category would support the hypothesis "There are multiple gods". In short, the BDMR posits a fundamental role for prior beliefs in religious reasoning and identifies several types of such beliefs including religious, interpersonal and social beliefs.

Note that in the BDMR prior beliefs can be treated as context-dependent (a similar approach is proposed by Taves & Asprem, 2017). In other words, in one context (e.g., a particular place, time, or with particular interlocutors) one claim might be assigned higher prior probability than another claim, while in a different context the opposite might occur. Take an individual entertaining prior beliefs regarding Christian faith against atheism. In one context (e.g., when attending the Christmas eve mass), prior beliefs supporting Christian faith might prevail over atheism, while in a different context (e.g., at the disco during new year's eve), prior beliefs supporting Christianity might be overshadowed by atheism. Because of such context-dependent nature of prior beliefs, within the BDMR reasoning will also be context-dependent, meaning that a hypothesis will be more likely to be accepted when consistent with prior beliefs. The context-dependent nature of religious beliefs is now supported by substantial evidence (Bialecki, 2017; Legare & Visala, 2011; Legare et al., 2012; Luhrmann, 2018; Shtulman & Legare, 2019; Shtulman & Lombrozo, 2016). For example, in many domains explanatory

coexistence has been observed, occurring when people alternate between natural and supernatural explanations of phenomena like illness (Legare & Visala, 2011; Legare et al., 2012; Shtulman & Legare, 2019; Shtulman & Lombrozo, 2016). Anthropological investigations have also documented religious and mundane attitudes alternating within the very same persons, a phenomenon described by distinguishing between a faith frame (guiding behaviour in holy contexts) and an everyday frame (guiding behaviour in mundane contexts) (Luhrmann, 2018).

In addition to prior beliefs, the BDMR proposes that available evidence impacts substantially on religious reasoning (Fig 2B). Two types of evidence are postulated by the model, one pertaining direct observations (DirE) and the other concerning information gathered from social sources (SocE). Regarding the latter, it is evident how social sources such as sacred texts and priests are often influential in the formation of religious beliefs. A similar influence can be ascribed to the own family and community, as well as to media such as magazines and television. For example, it has been observed that people are more likely to endorse religious views handed down by their parents (Granqvist et al., 2007; Stark & Finke, 2000). In the BDMR, information provided by social sources is implemented by the variable SocE. Consider the case above arbitrating between the hypothesis that an illness is God's punishment and the hypothesis that the illness is due to a frequent interaction with an infected patient. For example, a family member or a priest might express an opinion supporting the first hypothesis. This opinion is represented by the BDMR as a SocE observation and is considered during reasoning. Importantly, the model attributes a weight to this information which determines how influential this will be during reasoning. For example, family members might be considered as ignorant on religious matters, and hence their opinion might be considered only marginally. On the contrary, the priest's opinion might be considered as highly reliable and hence weighted heavily during reasoning.

The second type of evidence implemented by the BDMR concerns direct observation acquired via own senses (DirE). For example, one might believe that certain observations (e.g., pain which increases

when walking next to the church) fit better with the hypothesis that an illness is God's punishment, and that other observations (e.g., viewing other people getting sick after interacting with infected patients) fit better with the hypothesis that the illness is due to a frequent interaction with an infected patient. As for SocE, direct observation is also associated with a weight which determines its relevance. Empirical evidence indicates that certain experiences such as dreams, ecstatic states, and visions are often treated as particularly relevant during religious reasoning (Bulkeley, 2016), possibly because these experiences are difficult to understand in terms of everyday life. For example, one might consider dreams as particularly revealing and hence weight them heavily during religious reasoning.

By relying on prior beliefs (PBS) and observations (SocE and DirE), the BDMR offers an interpretation of explanation theories of religion expressed in formal computational terms. These theories suggest that the primary drive for religion is to explain salient aspects of life and reality. The BDMR captures this perspective by proposing that explanations can be derived by integrating prior beliefs and novel evidence. Prior knowledge appears in three forms, as religious beliefs available a priori, as expectations about interpersonal relationships, and as knowledge about the structure of society. Novel evidence can concern information gathered from social actors such as people, media, and institutions, as well as from the own perception. In the next section, I will examine one last key element of the model, namely the utility component, and I will analyse its connection with motivation theories of religion.

## 4. Utility and motivation theories

Although prior beliefs and observations are fundamental elements of the BDMR, they are unable to account for aspects of religion highlighted by motivation theories. The latter propose that religious beliefs ultimately do not arise from an attempt to explain, but from other motives such as promoting community bonds, exerting power, reducing anxiety, or supporting moral rules. The BDMR takes these motives into account by including a utility component implemented by the variable EOut (Fig. 3).

Consider the example above arbitrating between the claim that an illness is God's punishment for recent misbehaviour (a religious hypothesis), and the alternative statement that an illness is due to a frequent interaction with an infected patient (a non-religious hypothesis). By relying on EOut, the model asks: what is the consequence (in terms of utility) of accepting the religious hypothesis (and spending time praying) if this is true? And if it not true? What is the consequence of accepting the non-religious hypothesis (and not spending time praying) if this is true? And if it is not true? Based on the answer to these questions (and on estimating how likely each hypothesis is based on prior knowledge (PBS) and evidence (DirE and SocE)), the model eventually accepts one hypothesis as true. In other words, in addition to considering the likelihood of the hypotheses, the model postulates that considerations about their utility have an immediate impact upon reasoning. Phenomenologically, the proposal is that individuals are blind to these considerations about utility, and think that their beliefs are the result of factual considerations only.

As another example of the role of utility in the model, consider the hypothesis "kings are chosen by God" versus "kings happen to rule by chance". Explanation theories propose that a medieval king would ponder the different evidence in favour or against each hypothesis and select the most plausible thereof. Conversely, motivation theories argue that, independent of any evidence, that king would be biased towards the first hypothesis because this promotes his own personal interests; such bias would act unconsciously. Following motivation theories, the BDMR explains the bias favouring the first hypothesis as deriving from the questions: what is the consequence (in terms of utility) of accepting the hypothesis "kings are chosen by God" if this is true? And if it not true? What is the consequence of accepting the hypothesis "kings happen to rule by chance" if this is true? And if it is not true? Answering these questions would lead to a bias for the first hypothesis: its rejection would appear as costlier (in terms of own interest) than rejecting the alternative. In line with motivation theories, this reasoning process is postulated to be unconscious.

The utility associated with rejection and acceptance of each hypothesis depends on the specific value attributed to the possible outcomes. The BDMR proposes that different utility values can be attributed by different individuals, and that the same individual might attribute different values in different contexts (Kahoe, 1985; Saroglou, 2011) (why certain utility values are attributed to outcomes is not the focus of the model). Again, consider the example above comparing a religious versus non-religious hypothesis. One individual facing this dilemma might not be frightened at all by the illness. Such indifference would imply a large cost if the religious hypothesis is accepted (and time is spent praying for receiving God's help to heal - under the assumption that time spent praying is costly) but the hypothesis turns out to be false (and hence praying turns out to be useless). Hence, this individual will be likely to accept the non-religious hypothesis. On the contrary, another individual facing the same dilemma might be extremely frightened by the illness. For this person, a large cost occurs if the religious hypothesis is rejected (and prayer is not performed) but the hypothesis turns out to be true (and hence God's help to heal is not received). The second individual will be likely to accept the religious hypothesis.

The value attributed to outcomes might sometimes depend on a unique motivation, such as promoting community bonding, and other times on multiple motivations that need to be integrated, such as both promoting bonding and increasing the own power (Saroglou, 2011). The relative importance of each motive might differ across individuals and across contexts, for example with some people favouring bonding promotion over power increase, and other people the other way around.

The inclusion of the utility component enables the model to interpret motivation theories, and to reconcile them under a unifying framework. It is well established that a central human drive is to promote social bonds and a sense of belonging (Baumeister, 1991; Baumeister & Leary, 1995). Empirical data support the possibility that this drive is critical in religion (Baumeister, 1991; Krause & Wulff, 2005; Pargament et al., 1983). Within the BDMR, this notion can be implemented by EOut in such a way that a religious hypothesis will be more likely to be endorsed when its acceptance is viewed

as facilitating bonds and sense of belonging. This notion is analogous to motivation theories which explain religion as deriving from encouraging a sense of group belonging and group bonding (Baumeister, 1991; Krause & Wulff, 2005; Pargament et al., 1983). A second fundamental human drive is to increase the status, power, and economic wealth of the self and of the own group (sometimes at the expense of other people and other groups) (Jost et al., 2014; Williams, 1996; Fanon, 1963; Fulton, 1987). This drive can also be implemented by EOut, implying that a religious hypothesis will be more likely to be endorsed when its acceptance is viewed as promoting the status, power, and economic wealth of the self and of the own group. This reasoning is analogous to motivation theories which explain religious beliefs as an ideology embraced by a social class or group to promote its interest (Jost et al., 2014; Williams, 1996; Fanon, 1963; Fulton, 1987). A third important human motivation consists in promoting moral behaviour and justice (Turiel, 2002). Such motivation has been linked with religion (Bloom, 2012; McKay & Whitehouse, 2015), in line with the finding of a positive association between religiosity and willingness to be a moral and virtuous individual (Batson et al., 1993; Vitell et al., 2009). A motivation for promoting morality and justice can be captured by EOut in such a way that a hypothesis will be more likely to be endorsed when its acceptance is viewed as supporting morality and justice. This reasoning is analogous to motivation theories which explain religious beliefs as reflecting an effort to promote morality and justice (Batson et al., 1993; Saroglou, 2011; Vitell et al., 2009).

One last group of motivation theories of religion (comfort theories) proposes that religious beliefs are embraced because they suppress fear and anxiety (especially death anxiety) (Atran & Norenzayan, 2004; Norenzayan & Hansen, 2006; Swan, 2019; Willer, 2009; Vail et al., 2010). This suppression would occur because religious beliefs would call upon positive interpretations of life and reality and dismiss negative interpretations. For example, in this perspective belief in an after-life is proposed to emerge because it suppresses the thought that spiritual life ends with material death, a thought which is assumed to elicit anxiety. The BDMR implicates an alternative perspective, because according to this model the belief selected is not necessarily the most positive interpretation but, crucially, the one

more costly to reject. As an example, consider someone highly scared by an illness and arbitrating between the hypothesis that the illness will not be contracted if God is prayed, but it will be contracted otherwise, versus the hypothesis that the illness will not be contracted anyway. Comfort theories would predict that (other things being equal) the second hypothesis will be endorsed, because it is the most positive, and hence the one which best suppresses anxiety. Conversely, the BDMR postulates that the first hypothesis will be endorsed, because, although it is more anxiety-provoking, it is the costliest to reject. If this hypothesis is rejected and God is not prayed, the risk of contracting the illness increases. I argue that the perspective offered by the BDMR fits better with the fact that religious beliefs often appear to enhance, rather than inhibit, negative emotions (Boyer, 2001; Atran & Norenzayan, 2004). For example, there are cultures (e.g., the ancient Greeks and Mesopotamians) where gods and spirits have prevalently an ambivalent, or even negative, attitude towards humans (Boyer, 2001). Also, some religions do not focus on the afterlife, or have a rather gloomy view of it (Boyer, 2001). These beliefs appear as fuelling, rather than suppressing, anxiety. However, note that the BDMR predicts that emotions such as anxiety influence religious reasoning, albeit not in the way suggested by comfort theories. This prediction occurs because emotions have an impact on the utility values attributed to the different outcomes, hence affecting religious reasoning. The research on the role of emotions on religious reasoning has focused on anxiety (especially regarding death) and has produced mixed findings (Jong & Halberstadt, 2017; Jong et al., 2018), although in general it suggests that anxiety is influential. Within this literature, an observation which appears as robust is that feeling a loss of control over the environment favours religious over non-religious beliefs (Kay et al., 2008; 2010; McGregor et al., 2010; Whitson & Galinsky, 2008). This observation is compatible with the BDMR, as described in the next section in details.

In short, the inclusion of a utility component is a critical feature of the BDMR which allows the model to interpret motivation theories of religion. The idea is that beliefs are more likely to be endorsed when they encourage motives such as fostering group bonding and sense of belonging, supporting the status and power of the self and the own group, and promoting morality. Importantly, these motives

can be fulfilled only if an individual appraises reality with some degree of accuracy. For example, a careful examination of reality is necessary to identify which beliefs are better suited for supporting group bonding. The requirement of appraising reality accurately is often disregarded by motivation theories, but it is emphasised by the BDMR (and by explanation theories) thanks to the role played by prior beliefs and evidence that I have discussed above.

## 5. Empirical implications

One of the central aims of this paper is to introduce a framework that can help interpreting empirical observations and generating new predictions. To this aim, adopting a computational approach has several advantages in as much as it relies on a precise and formal description. To illustrate how the BDMR can be adopted to interpret empirical phenomena, this section examines the model in the context of recent empirical research about the psychology of religion.

Substantial evidence has shown that humans exhibit an innate propensity to interpret certain sensory experiences in terms of agency (Atran, 2002; Bloom, 2007; Barrett, 2000; Guthrie, 1993). This propensity appears early in childhood and has been observed among different cultures (Barrett, 2000; Guthrie, 1993). These observations have inspired the proposal that a bias towards agency-detection is at the root of religion's emphasis on supernatural agents such as gods and spirits (Atran, 2002; Bloom, 2007; Barrett, 2000; Guthrie, 1993). However, there is poor evidence that an agency-detection bias plays a causal role in the formation of religious beliefs (Van Elk & Van Leeuwen, 2019). In light of this, there is still debate on the precise role of agency-detection in religion (Van Elk & Van Leeuwen, 2019). According to a recent proposal (Van Elk & Van Leeuwen, 2019), religious beliefs can be distinguished in general and personal beliefs, where the former concern broad statements transmitted by the own culture (e.g., "God speaks with people having true faith"), and the latter concern interpretations of direct experience (e.g., "This morning God talked to me while I was praying"). When exposure to certain conditions elicits an agency experience (i.e., the perception that

some agent is present), this would be interpreted by relying on general beliefs, eventually leading to the formation of personal beliefs (Van Elk & Van Leeuwen, 2019). For example, someone believing that "God reveals himself to people having true faith", and who has an agency experience during the morning prayer, might interpret the experience as "This morning God joined me while I was praying". This proposal fits with the BDMR. General beliefs can be represented by PBS; for example, a person might consider two alternative prior beliefs: "God reveals himself to people having true faith" versus "God never reveals himself to individuals". Personal beliefs can be represented by Hyp: for example, the person might evaluate two possible hypotheses: "This morning God joined me while I was praying" versus "This morning my uncle briefly entered in the room while I was praying". The model would also assume that the hypothesis "This morning God joined me while I was praying" is more likely if the statement "God reveals himself to people having true faith" is true. Finally, an agency experience can be represented by DirE. What happens when an agency experience occurs? The model predicts that the consequence depends on the prior probability associated with the two prior beliefs (implemented in PBS). Someone assigning higher probability to the statement "God reveals himself to people having true faith" will tend to believe that "This morning God joined me while I was praying" after the agency experience occurs. This example shows that the BDMR proposes a role for agency detection which is similar to recent proposals (Van Elk & Van Leeuwen, 2019).

The way the BDMR interprets agency detection is analogous to explanations offered by predictive coding (Andersen, 2019; Schjoedt et al., 2013; Taves & Asprem, 2017; Van Elk et al., 2016). The latter has been recently proposed to interpret important processes underlying the psychology of religion, such as how prior cultural knowledge drives formation of religious beliefs (Taves & Asprem, 2017). PBS, Hyp, DirE and SocE allow the BDMR to implement Bayesian inference, which also underlies predictive coding. Hence, phenomena that can be fruitfully described by predictive coding can be captured equally well by the BDMR. However, by including the utility component, the BDMR offers a way to go beyond predictive coding and explain phenomena where apparently puzzling emotional and motivational processes are at play. As an example, consider research showing that feeling lack of

control enhances religious faith (Kay et al., 2008; 2010). In a study, participants were asked to remember a recent positive event during which they felt having no control (Kay et al., 2008). After remembering the event, participants were more likely to report enhanced belief in God. Predictive coding does not seem to offer much insight on this effect. Appealing to comfort theories, some have interpreted this effect as occurring because lack of control would elicit anxiety, and because believing in God would suppress such anxiety (Kay et al., 2010). However, against this explanation, in the experiment the event to remember was positive, and consistently participants reported no change in affect. This casts doubts on whether anxiety played any role. An alternative interpretation is offered by the BDMR. This proposes that the experimental paradigm (requiring to remember an event characterised by lack of control and subsequently to report the strength of the belief in God) might induce participants to wonder what forces drive their everyday life, and to consider three alternative hypotheses: the self, God, or uncontrollable external forces. Focusing on a poorly controllable event (evoked during the experiment) would support the God hypothesis and the uncontrollable-external-forces hypothesis at the expense of the self hypothesis. But how to arbitrate between the God and the uncontrollable-external-forces hypothesis? Here the BDMR proposes that a key role is played by the utility component, implemented by EOut. The participant would ask: what happens (in terms of utility) if the God hypothesis is true and I accept it (and, say, respect God's commandments)? And if I reject it (and ignore God's commandments)? What happens if the uncontrollable-external-forces hypothesis is true and I accept it? And if I reject it? For most people, answering these questions might favour the God hypothesis over the uncontrollable-external-forces hypothesis. This is because rejecting the God hypothesis (and ignore God's commandments) if this turns out to be true would appear as highly costly: God favour will not be won. Moreover, accepting the uncontrollable-external-forces hypothesis does not seem very appealing, even if this turns out to be true: by definition, uncontrollable external forces remain impossible to discipline. In other words, the BDMR proposes that the effects found in these experiments (Kay et al., 2008; 2010) might occur because rejecting the

possibility of God's existence is perceived to be costlier than rejecting the possibility that life depends on uncontrollable forces.

In short, I have discussed two examples of how the BDMR can be adopted to interpret empirical phenomena. The model subsumes predictive coding, meaning that in some cases the two can be used interchangeably. However, by including a utility component, the model goes beyond predictive coding and offers explanations for cases where motivational and emotional factors influence religious beliefs.

## 6. Discussion

This paper introduces the BDMR, which offers a computational framework for reconciling explanation and motivation theories of religion. Key elements of the BDMR are prior beliefs and (direct and social) evidence (which allows the model to describe explanation theories) and expected utility (which allows the model to describe motivation theories). It is important to highlight that the BDMR operates at a computational level, in other words it seeks to identify the first principles, or the basic meaning, of a psychological phenomenon (though this level of analysis speaks also to the implementation (e.g., neural) level, as both will need to be ultimately integrated together). Such level of analysis presupposes that a psychological phenomenon is ultimately rational, in the sense that it is grounded on processes that are adaptive for survival and reproduction. A different level of analysis focuses on the fine-grained mechanisms underlying religious reasoning. At this level, empirical research has found several biases and illogicities (Shermer, 2002; Wolpert, 1994). This has led some scholars to consider religious reasoning as fundamentally irrational, implying that exploring its computational principles can be considered as futile (Wolpert, 1994). However, two considerations can be made against this argument. First, similar biases and illogicities are observed also when reasoning does not involve any religious hypothesis (Shermer, 2002). Hence, it remains dubious whether in everyday life reasoning becomes more irrational when it comprises religious content. Second, both religious and non-religious reasoning can be viewed as guided by bounded rationality (Stark & Finke, 2000). In this

view, biases and illogicities, interpreted as cognitive shortcuts adopted by the brain because of its limited computational capacity, still produce satisfying results. Linked to this issue is the question of whether religious and non-religious reasoning are driven by similar or different processes. The perspective offered by the BDMR suggests that similar processes might be engaged.

This paper has emphasised the connection between the BDMR and previous theories of religion, both explanation and motivation accounts. The frameworks with the strongest analogies with the BDMR are the predictive coding (Andersen, 2019; Schjoedt et al., 2013; Taves & Asprem, 2017; Van Elk et al., 2016) and rational choice (Iannaccone, 1998; Stark & Finke, 2000) approach; so much so that the BDMR can be conceived as a synthesis of the two. The BDMR and predictive coding are substantially equivalent when examining explanation theories. However, predictive coding is insufficient to account for motivation theories. To address this, the BDMR extends predictive coding to Bayesian decision (Bishop, 2006) and utility maximization. Notably, predictive coding has been recently generalised to explain emotional and motivational processes (e.g., Friston et al., 2015; Seth, 2013). An interesting question for future research is whether this revised formulation of predictive coding can integrate explanation and motivation theories in a way analogous to the BDMR.

The notion that utility maximization is pivotal in religion is central to the BDMR as much as it is to the rational choice model (Iannaccone, 1998; Stark & Finke, 2000). However, I highlight two fundamental differences between the BDMR and the rational choice model. First, the BDMR (similar to predictive coding) describes explicitly how individuals consider prior beliefs and evidence in the formation of their beliefs, while the rational choice model does not focus on this. Second, and more fundamentally, the rational choice model can be ultimately considered an explanation theory because it presupposes that utility does not affect religious beliefs. In other words, the religious beliefs are conceived as attempts to describe life and reality accurately. Utility does not have any influence at the time when beliefs are formed, coming into play only later when these beliefs are considered to make choice. Coherent with this assumption, proponents of the rational choice model have criticised motivation

theories because the latter would presuppose a role for motivation in affecting religious beliefs (Stark & Finke, 2000). On the contrary, the BMDR proposes that utility comes into play immediately during religious reasoning, implying that religious beliefs are not the consequence of a pure attempt to explain, but also of other motives highlighted by motivation theories. Notably, this does not affect the general idea that religious reasoning is rational, although now in pragmatic, rather than epistemic, terms.

The model focuses on a specific aspect of the psychology of religion: belief formation. It is important to integrate this within a broader picture which aims at explaining religious behaviour. In some cases, the link between belief and behaviour might be direct and automatic (e.g., consider someone praying everyday under the conviction that this is necessary for salvation). Other times, behaviour might emerge from deliberative processes based on weighting costs and benefits of different courses of action (e.g., some scholars have proposed that careful deliberation often underlies religious conversion; Lofland & Stark, 1965; Long & Hadden , 1983). An interesting avenue for future research is to adopt computational modelling for exploring the processes linking religious beliefs and behaviour. With this regard, a common assumption is that behaviour is the *consequence* of beliefs. However, some scholars have argued that religious behaviour is often the *cause* of religious beliefs (e.g., Argyle, 2006). In other words, individuals might first participate in religious activities and then provide post-hoc justification for their participation. This phenomenon can be explored adopting the BDMR: this model is suitable to investigate post-hoc rationalizations following performance of religious behaviour.

In short, I introduce a model of religious reasoning which examines the underlying computational principles. The model has an integrative scope as it attempts to reconcile different, and often competing, perspectives, acknowledging that each perspective sheds light on essential aspects of religion. The model can provide a conceptual framework for further theoretical and empirical investigations. Moreover, because the notion of Bayesian decision (on which the BDMR is built upon)

represents a promising general framework, the model can bridge research on religious reasoning with research on other forms of reasoning.

## 6. Appendix

Consider the example above arbitrating between the claim that an illness is God's punishment for recent misbehaviour (a religious hypothesis), and the claim that an illness is due to a frequent interaction with an infected patient (a non-religious hypothesis). Prior beliefs are that God often intervenes in people's life versus the alternative view that God is usually uninterested in mundane affairs. Formally, the model is a mixture of Gaussians. The joint probability can be written as:

P(PBS, Hyp, HDec, EOut, DirE, SocE) = P(PBS) P(HDec) P(Hyp|PBS) P(DirE|Hyp) P(SocE|Hyp) P(EOut|Hyp, HDec)

PBS is a categorical variable with number of categories equal to $n_{PBS}$ and where each category is associated with a probability. In the example, one can set $n_{PBS} = 2$, *PBS = Int* if God is interventionist, and *PBS = NoInt* if God is not interventionist. The probability of God being interventionist is *P(PBS = Int) = x* and the probability of God not being interventionist is *P(PBS = NoInt) = 1 - x* (where $0 \leq x \leq 1$). Hyp is also categorical, with number of categories equal to $n_{Hyp}$. Considering the example, one can set $n_{Hyp} = 2$, *Hyp = Rel* for the religious hypothesis (the illness is God's punishment), and *Hyp = NoRel* for the non-religious hypothesis (the illness is due to interacting with an infected patient). The conditional probabilities for Hyp are *P(Hyp = Rel | PBS = Int) = y*, *P(Hyp = NoRel | PBS = Int) = 1 – y*, *P(Hyp = Rel | PBS = NoInt) = z*, *P(Hyp = NoRel | PBS = NoInt) = 1 – z* (where $0 \leq y \leq 1$ and $0 \leq z \leq 1$). HDec is also categorical, with the number of categories $n_{HDec} = n_{Hyp}$. In the example, *HDec = RelAcc* when the religious hypothesis is accepted (or, equivalently, when the non-religious hypothesis is rejected) and *HDec = NoRelAcc* when the religious hypothesis is rejected (or, equivalently, when the non-religious hypothesis is accepted). Probabilities for HDec are *P(HDec = RelAcc) = u* and *P(HDec = NoRelAcc) = 1 - u* (where $0 \leq u \leq 1$).

DirE is a Gaussian variable conditioned on Hyp. Its conditional probability can be defined as:

$$P(\text{DirE I Hyp = k}) = \mathcal{N}(\mu_{DirE|k}, 1/{\lambda_{DirE}}^2)$$

Here, every category of Hyp *k* has its own associated average $\mu_{DirE|k}$; for instance the model will include $\mu_{DirE|Rel}$ (conditional on the religious hypothesis being true) which is different from $\mu_{DirE|NoRel}$ (conditional on the non-religious hypothesis being true). The parameter ${\lambda_{DirE}}^2$ reflects the weight or precision of DirE and in the model it is equal for all levels of Hyp (in principle, a specific weight for each level of Hyp can be implemented). A similar logic applies to SocE, where the conditional probability is:

$$P(\text{SocE I Hyp = k}) = \mathcal{N}(\mu_{SocE|k}, 1/{\lambda_{SocE}}^2)$$

Also for SocE, every category of Hyp *k* has its own associated average $\mu_{SocE|k}$; for instance the model will include $\mu_{SocE|Rel}$ (conditional on the religious hypothesis being true) which is different from $\mu_{SocE|NoRel}$ (conditional on the non-religious hypothesis being true). The parameter ${\lambda_{SocE}}^2$ reflects the weight or precision of SocE and in the model it is equal for all levels of Hyp (in principle, a specific weight for each level of Hyp can be implemented).

Finally, EOut is a Gaussian variable conditioned on both Hyp and HDec. Its conditional probability is:

$$P(\text{EOut I Hyp = k, HDec = j}) = \mathcal{N}(\mu_{EOut|k,j}, \sigma^2_{Eout})$$

This indicates a specific average exists for each combination of Hyp and HDec. For instance, the model comprises $\mu_{EOut|Rel,RelAcc}$ (the expected outcome if the religious hypothesis is true and it is correctly accepted), $\mu_{EOut|Rel,NoRelAcc}$ (the expected outcome if the religious hypothesis is true but it is wrongly rejected), $\mu_{EOut|NoRel,NoRelAcc}$ (the expected outcome if the non-religious hypothesis is true and it is correctly accepted), $\mu_{EOut|NoRel,RelAcc}$ (the expected outcome if the non-religious hypothesis is true but it is wrongly rejected). The parameter $\sigma^2_{Eout}$ reflects the uncertainty about the outcome and in the

model it is equal for all combinations of Hyp and HDec (although in principle one can also implement a specific weight for each combination).

The model is used to make inference. For inference, the variables DirE and SocE are observed, while the other variables are not. The inference process includes multiple inference steps. At each step, for each level of HDec *j*, the model infers the conditional probability of EOut given the observed values for DirE and SocE and given HDec = *j*. This corresponds to the posterior Gaussian distribution:

$$P(EOut|\ DirE\ ,\ SocE,\ HDec = j) = \mathcal{N}(\mu_{EOut|DirE,SocE,j}, \sigma^2_{POST})$$

Where $\mu_{EOut|DirE,SocE,j}$ is the posterior average for the expected outcome (the parameter $\sigma^2_{POST}$ reflects the posterior uncertainty). For example, $\mu_{EOut|DirE,SocE,RelAcc}$ will be the posterior average if the religious hypothesis is accepted and $\mu_{EOut|DirE,SocE,NoRelAcc}$ is the posterior average if the non-religious hypothesis is accepted.

After these inferences are made, the model makes a decision by choosing the hypothesis associated with the highest posterior $\mu_{EOut|DirE,SocE,j}$. For instance, it will either choose to accept or reject the religious hypothesis (or, equivalently, to reject or accept the non-religious hypothesis, respectively).

**References**


Abelson, R. P. (1981). Psychological status of the script concept. *American psychologist*, *36*(7), 715.

Andersen, M. (2019). Predictive coding in agency detection. *Religion, Brain & Behavior*, *9*(1), 65-84.

Argyle, M. (2006). *Religious behaviour*. Routledge.

Atran, S. (2002). *In gods we trust: The evolutionary landscape of religion*. Oxford University Press.

Atran, S., & Norenzayan, A. (2004). Religion's evolutionary landscape: Counterintuition, commitment, compassion, communion. *Behavioral and brain sciences*, *27*(6), 713-730.

Barrett, J. L. (2000). Exploring the natural foundations of religion. *Trends in cognitive sciences*, *4*(1), 29-34.

Batson, C. D., Schoenrade, P., & Ventis, W. L. (1993). *Religion and the individual: A social-psychological perspective*. Oxford University Press.

Baumeister, R. F. (1991). Meanings of life New York: Guilford.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological bulletin*, *117*(3), 497.

Bialecki, J. (2017). *A Diagram for Fire: miracles and variation in an American charismatic movement* (Vol. 21). Univ of California Press.

Bloom, P. (2012). Religion, morality, evolution. *Annual review of psychology*, *63*, 179-199.

Bloom, P. (2007). Religion is natural. *Developmental science*, *10*(1), 147-151.

Boyer, P. Religion explained: The evolutionary origins of religious thought 2001 New York. *NY: Basic Books*.

Bowlby, J. (1969). *Attachment*. New York: Basic Books

Bulkeley, K. (2016). *Big dreams: The science of dreaming and the origins of religion*. Oxford University Press.

Cassibba, R., Granqvist, P., Costantini, A., & Gatto, S. (2008). Attachment and god representations among lay Catholics, priests, and religious: A matched comparison study based on the adult attachment interview. *Developmental Psychology*, *44*(6), 1753.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, *36*(3), 181-204.

Durkheim, E. (1912). *The elementary forms of the religious life.*

Iannaccone, L. R. (1998). Introduction to the Economics of Religion. *Journal of economic literature*, *36*(3), 1465-1495.

Evans-Pritchard, E. E. (1937). *Witchcraft, oracles and magic among the Azande* (Vol. 12). London: Oxford.

Fanon (1963). *The wretched of the earth* (Vol. 36). New York: Grove Press.

Freud, S. (1927). *The future of an illusion*. Broadview Press.

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1211-1221.

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive neuroscience*, *6*(4), 187-214.

Gramsci, A. (1971). *Selections from the prison notebooks*. London: Lawrence and Wishart.

Granqvist, P., Ivarsson, T., Broberg, A. G., & Hagekull, B. (2007). Examining relations among attachment, religiosity, and new age spirituality using the Adult Attachment Interview. *Developmental Psychology*, *43*(3), 590.

Granqvist, P., Mikulincer, M., & Shaver, P. R. (2010). Religion as attachment: Normative processes and individual differences. *Personality and Social Psychology Review*, *14*(1), 49-59.

Guthrie, S. E. (1993). *Faces in the clouds: A new theory of religion*. Oxford University Press on Demand.

Hogg, M. A., Adelman, J. R., & Blagg, R. D. (2010). Religion in the face of uncertainty: An uncertainty-identity theory account of religiousness. *Personality and social psychology review*, *14*(1), 72-83.

Hume, D. (1757). The Natural History of Religion, London: A. and H. *Bradlaugh Bonner*.

Jost, J. T., Hawkins, C. B., Nosek, B. A., Hennes, E. P., Stern, C., Gosling, S. D., & Graham, J. (2014). Belief in a just God (and a just society): A system justification perspective on religious ideology. *Journal of Theoretical and Philosophical Psychology*, *34*(1), 56.

Jong, J., & Halberstadt, J. (2017). What is the causal relationship between death anxiety and religious belief?. *Religion, Brain & Behavior*, *7*(4), 296-298.

Jong, J., Ross, R., Philip, T., Chang, S. H., Simons, N., & Halberstadt, J. (2018). The religious correlates of death anxiety: A systematic review and meta-analysis. *Religion, Brain & Behavior*, *8*(1), 4-20.

Kahoe, R. D. (1985). The development of intrinsic and extrinsic religious orientations. *Journal for the Scientific Study of Religion*, *24*(4), 408-412.

Kay, A. C., Gaucher, D., Napier, J. L., Callan, M. J., & Laurin, K. (2008). God and the government: testing a compensatory control mechanism for the support of external systems. *Journal of personality and social psychology*, *95*(1), 18.

Kay, A. C., Gaucher, D., McGregor, I., & Nash, K. (2010). Religious belief as compensatory control. *Personality and Social Psychology Review*, *14*(1), 37-48.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, *27*(12), 712-719.

Krause, N., & Wulff, K. M. (2005). " Church-Based Social Ties, A Sense of Belonging in a Congregation, and Physical Health Status". *The International Journal for the Psychology of Religion*, *15*(1), 73-93.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin*, *108*(3), 480.

Legare, C. H., & Visala, A. (2011). Between religion and science: Integrating psychological and philosophical accounts of explanatory coexistence. *Human Development*, *54*(3), 169-184.

Legare, C. H., Evans, E. M., Rosengren, K. S., & Harris, P. L. (2012). The coexistence of natural and supernatural explanations across cultures and development. *Child development*, *83*(3), 779-793.

Lofland, J., & Stark, R. (1965). Becoming a world-saver: A theory of conversion to a deviant perspective. *American sociological review*, 862-875.

Long, T. E., & Hadden, J. K. (1983). Religious conversion and the concept of socialization: Integrating the brainwashing and drift models. *Journal for the Scientific Study of Religion*, 1-14.

Luhrmann, T. M. (2018). The faith frame: Or, belief is easy, faith is hard. *contemporary pragmatism*, *15*(3), 302-318.

Marx, K., & Engels, F. (1845). The German Ideology.

McCauley, R. N. (2017). Twenty-five years in: Landmark empirical findings in the cognitive science of religion. *Religion Explained?: The Cognitive Science of Religion after Twenty-five Years*, 17.

McGregor, I., Nash, K., & Prentice, M. (2010). Reactive approach motivation (RAM) for religion. *Journal of Personality and Social Psychology*, *99*(1), 148.

McKay, R., & Whitehouse, H. (2015). Religion and morality. *Psychological bulletin*, *141*(2), 447.

Norenzayan, A., & Hansen, I. G. (2006). Belief in supernatural agents in the face of death. *Personality and Social Psychology Bulletin*, *32*(2), 174-187.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press.

Pargament, K. I., Silverman, W., Johnson, S., Echemendia, R., & Snyder, S. (1983). The psychosocial climate of religious congregations. *American Journal of Community Psychology*, *11*(4), 351-381.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, *2*(1), 79-87.

Saroglou, V. (2011). Believing, bonding, behaving, and belonging: The big four religious dimensions and cultural variation. *Journal of Cross-Cultural Psychology*, *42*(8), 1320-1340.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in cognitive sciences*, *17*(11), 565-573.

Schjoedt, U., Sørensen, J., Nielbo, K. L., Xygalatas, D., Mitkidis, P., & Bulbulia, J. (2013). Cognitive resource depletion in religious interactions. *Religion, Brain & Behavior*, *3*(1), 39-55.

Schwardmann, P., & Van der Weele, J. (2019). Deception and self-deception. *Nature human behaviour*, *3*(10), 1055-1061.

Shermer, M. (2002). *Why people believe weird things: Pseudoscience, superstition, and other confusions of our time*. Macmillan.

Shtulman, A., & Legare, C. H. (2019). Competing Explanations of Competing Explanations: Accounting for Conflict Between Scientific and Folk Explanations. *Topics in cognitive science*.

Shtulman, A., & Lombrozo, T. (2016). Bundles of contradiction: A coexistence view of conceptual change. *Core knowledge and conceptual change*, 49-67.

Simon, H. A. (1997). *Models of bounded rationality: Empirically grounded economic reason* (Vol. 3). MIT press.

Smith, M. K., Trivers, R., & von Hippel, W. (2017). Self-deception facilitates interpersonal persuasion. *Journal of Economic Psychology*, *63*, 93-101.

Stark, R., & Finke, R. (2000). *Acts of faith: Explaining the human side of religion*. Univ of California Press.

Swan, T. P. D. (2019). *The Effect of Anxiety on Religious Cognition* (Doctoral dissertation, University of Otago).

Willer, R. (2009). No atheists in foxholes: Motivated reasoning and religious belief. *Social and psychological bases of ideology and system justification*, 241-264.

Taves, A., & Asprem, E. (2017). Experience as event: event cognition and the study of (religious) experiences. *Religion, Brain & Behavior*, *7*(1), 43-62.

Trivers, R. (2011). The folly of fools New York. *NY: Basic Books*.

Tylor, E. B. (1871). *Primitive culture: researches into the development of mythology, philosophy, religion, art, and custom* (Vol. 2). J. Murray.

Vail, K. E., Rothschild, Z. K., Weise, D. R., Solomon, S., Pyszczynski, T., & Greenberg, J. (2010). A terror management analysis of the psychological functions of religion. *Personality and Social Psychology Review*, *14*(1), 84-94.

van Elk, M., Friston, K., & Bekkering, H. (2016). The experience of coincidence: An integrated psychological and neurocognitive perspective. In *The challenge of chance* (pp. 171-185). Springer, Cham.

Van Leeuwen, N., & van Elk, M. (2019). Seeking the supernatural: The interactive religious experience model. *Religion, Brain & Behavior*, *9*(3), 221-251.

Vitell, S. J., Bing, M. N., Davison, H. K., Ammeter, A. P., Garner, B. L., & Novicevic, M. M. (2009). Religiosity and moral identity: The mediating role of self-control. *Journal of Business Ethics*, *88*(4), 601-613.

Whitson, J. A., & Galinsky, A. D. (2008). Lacking control increases illusory pattern perception. *science*, *322*(5898), 115-117.

Wolpert, L. (1994). *The unnatural nature of science*. Harvard University Press.

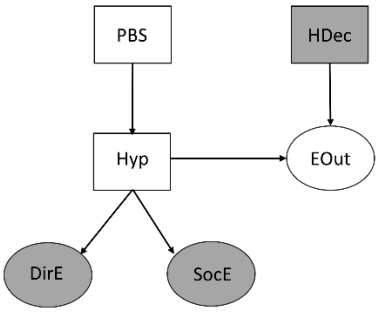


**Fig 1**. Bayesian network representing the model. Its variables are: Prior Belief Systems (PBS), Hypothesis (Hyp), Direct Evidence (DirE), Social Evidence (SocE), Hypothesis Decision (HDec), and Expected Outcome (EOut). Categorical and continuous variables are represented by rectangles and circles, respectively. Arrows indicate probabilistic causal relations from one variable to another. Shaded variables are those considered to be observed at each inference step.

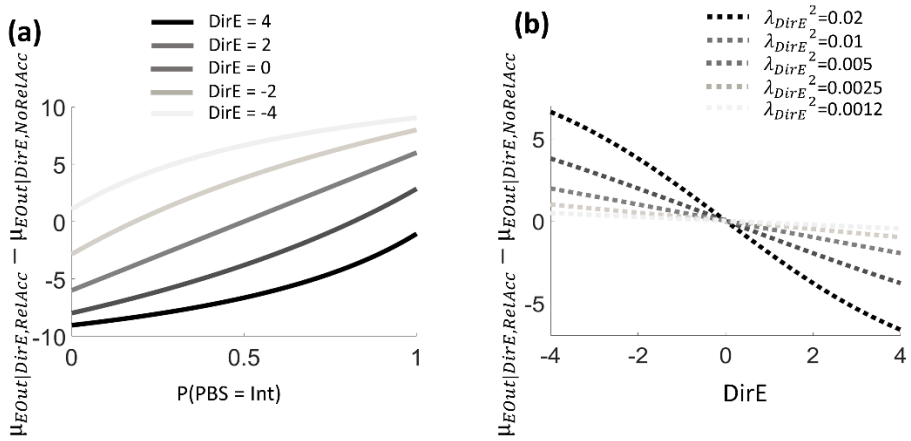


**Fig. 2**. Simulation of the model. The simulated scenario is discussed also in the main text and arbitrates between the claim that an illness is God's punishment for recent misbehaviour (a religious hypothesis) and the claim that an illness is due to a frequent interaction with an infected patient (a non-religious hypothesis). Hyp includes two categories (religious hypothesis vs non-religious hypothesis), PBS includes two categories (God is interventionist (Int) vs God is not interventionist (NoInt)), and negative values of DirE or SocE support the religious hypothesis. The y axis reflects the posterior outcome utility value of accepting the religious hypothesis minus the posterior outcome utility value of accepting the non-religious hypothesis. **A**: The x axis reflects the prior probability for PBS = Int. Different lines indicate different values for DirE (for all lines, SocE = 0, the precision parameter for DirE $\lambda_{DirE}{}^2 = 0.005$, the outcome of accepting the non-religious hypothesis when it is true ($\mu_{EOut|NoRel,NoRelAcc}$) is equal to zero, the outcome of accepting the non-religious hypothesis when it is false ($\mu_{EOut|Rel,NoRelAcc}$) is equal to -10, the outcome of accepting the religious hypothesis when it is true ($\mu_{EOut|Rel,RelAcc}$) is equal to zero, the outcome of accepting the religious hypothesis when it is false ($\mu_{EOut|NoRel,RelAcc}$) is equal to -10). **B**: The x axis reflects the value of DirE. Different lines indicate different values of the precision parameter for DirE $\lambda_{DirE}{}^2$ (for all lines, P(PBS = Int) = 0.5 and other parameters are as above). Note that an equivalent pattern would be produced if SocE was manipulated instead of DirE.

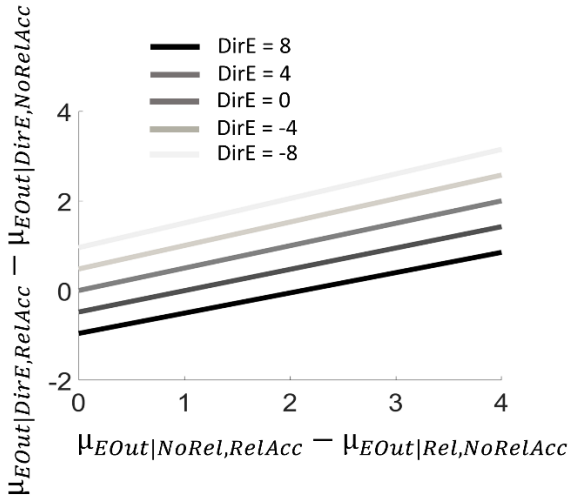


**Fig. 3**. Simulation of the model. The simulated scenario is discussed also in the main text and arbitrates between the claim that an illness is God's punishment for recent misbehaviour (a religious hypothesis) and the claim that an illness is due to a frequent interaction with an infected patient (a non-religious hypothesis). Hyp includes two categories (religious hypothesis vs non-religious hypothesis), PBS includes two categories (God is interventionist (Int) vs God is not interventionist (NoInt)), and negative values of DirE or SocE support the religious hypothesis. The y axis reflects the posterior outcome utility value of accepting the religious hypothesis minus the posterior outcome utility value of accepting the non-religious hypothesis. The x axis reflects the difference between the expected outcome of accepting the religious hypothesis when it is false ($\mu_{EOut|NoRel,RelAcc}$) and the expected outcome of accepting the non-religious hypothesis when it is false ($\mu_{EOut|Rel,NoRelAcc}$). Different lines indicate different values for DirE (for all lines, P(PBS = Int) = 0.5, SocE = 0, the precision parameter for DirE ${\lambda_{DirE}}^2 = 0.0012$, the expected outcome of accepting the non-religious hypothesis when it is true ($\mu_{EOut|NoRel,NoRelAcc}$) is equal to zero, the expected outcome of accepting the religious hypothesis when it is false ($\mu_{EOut|NoRel,RelAcc}$) is equal to -10, the expected outcome of accepting the religious hypothesis when it is true ($\mu_{EOut|Rel,RelAcc}$) is equal to zero).